Available online at www.sciencedirect.com
SCIENCE DIRECT
Advances in Mathematics 192 (2005) 310–340
ADVANCES IN Mathematics
http://www.elsevier.com/locate/aim

The Bellman functions of dyadic-like maximal operators and related inequalities

Antonios D. Melas*

Department of Mathematics, University of Athens, Room 105, GR-15784 Panepistimiopolis, Athens, Greece

Received 18 November 2003; accepted 22 April 2004

Communicated by C. Fefferman

Abstract

For each $p>1$ we precisely evaluate the main Bellman functions associated with the dyadic maximal operator on $\mathbb{R}^n$ and the dyadic Carleson imbedding theorem. Actually, we do that in the more general setting of tree-like maximal operators. These provide refinements of the sharp L^p inequalities for those operators. For this we introduce an effective linearization for such maximal operators on an adequate set of functions.

MSC: 42B25

Keywords: Dyadic; Maximal; Bellman

1. Introduction

The dyadic maximal operator on $\mathbb{R}^n$ is a useful tool in analysis and is defined by

$$M_d\phi(x) = \sup\left\{\frac{1}{|Q|}\int_Q |\phi(u)|\, du\colon\ x\in Q,\ \ Q\subseteq\mathbb{R}^n \text{ is a dyadic cube}\right\} \tag{1.1}$$

for every $\phi\in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ where the dyadic cubes are the cubes formed by the grids $2^{-N}\mathbb{Z}^n$ for $N=0,1,2,\ldots$.

*Fax: +30-210-727-6378.
E-mail address: amelas@math.uoa.gr.

0001-8708/$ - see front matter
doi:10.1016/j.aim.2004.04.013

As it is well known it satisfies the following weak type $(1,1)$ inequality

$$|\{x\in\mathbb{R}^n\colon M_d\phi(x)>\lambda\}|\leqslant\frac{1}{\lambda}\int_{\{M_d\phi>\lambda\}}|\phi(u)|\,du. \tag{1.2}$$

for every $\phi\in L^1(\mathbb{R}^n)$ and every $\lambda>0$ from which it is easy to get the following L^p inequality:

$$||M_d\phi||_p\leqslant\frac{p}{p-1}\,||\phi||_p \tag{1.3}$$

for every $p>1$ and every $\phi\in L^p(\mathbb{R}^n)$.

It is easy to see that the weak type inequality (1.2) is best possible. It has also been proved that (1.3) is best possible ([2,3] for the general martingales and [9] for dyadic ones).

In studying dyadic maximal operators as well as more general variants it would be convenient to work with functions supported in the unit cube $[0,1]^n$ and replace M_d by

$$M'_d\phi(x)=\sup\left\{\frac{1}{|Q|}\int_Q|\phi(u)|\,du: x\in Q, Q\subseteq[0,1]^n \text{ is a dyadic cube}\right\} \tag{1.4}$$

and hence work completely on the measure space $[0,1]^n$. A standard dilation and approximation argument allows one to pass to the operator M_d.

An approach for studying such maximal operators is the introduction of the so called Bellman functions (see [4]) related to them. These, aside from the fact that they allow easy proofs, reflect certain deeper properties of the maximal operators. We will be concerned with the following two ways of constructing such functions in relation to the maximal operators.

The first is defined for any $p>1$ by

$$\begin{aligned}&\mathcal{B}_p(F,f,L)\\ &\quad=\sup\left\{\frac{1}{|Q|}\int_Q(M_d\phi)^p:\ \mathrm{Av}_Q(\phi^p)=F,\mathrm{Av}_Q(\phi)=f,\ \sup_{R\,:\,Q\subseteq R}\mathrm{Av}_R(\phi)=L\right\},\end{aligned} \tag{1.5}$$

where Q is a fixed dyadic cube, R runs over all dyadic cubes containing Q, ϕ is nonnegative in $L^p(Q)$ and the variables F,f,L satisfy $0\leqslant f\leqslant L, f^p\leqslant F$. $\mathcal{B}_p$ is independent of the choice of Q (so we may take $Q=[0,1]^n$) and satisfies a certain "pseudoconvexity" inequality (see [4]). Constructing now a function satisfying the same properties provides good L^p bounds for the maximal operator M_d (see [4] for the details). For example the function $4F-4fL+2L^2$ is given in [4].

The second is related to the dyadic Carleson imbedding theorem and is defined for any $p>1$ by

$$\tilde{\mathcal{B}}_p(F,f,k)=\sup\left\{\frac{1}{|Q|}\sum_{R\subseteq Q}\lambda_R(\mathrm{Av}_R(\phi))^p:\ \mathrm{Av}_Q(\phi^p)=F,\mathrm{Av}_Q(\phi)=f,\right.$$
$$\left.\lambda_R\geqslant 0,\ \text{the } \lambda_R{}' \text{ satisfy condition } (C) \text{ and } \frac{1}{|Q|}\sum_{R\subseteq Q}\lambda_R=k\right\}, \tag{1.6}$$

where $0<k\leqslant 1$, $f^p\leqslant F$, Q is a fixed dyadic cube, R runs over all dyadic cubes containing Q, ϕ is nonnegative in $L^p(Q)$ and condition (C) means that for every dyadic $R\subseteq Q$ the following inequality is satisfied:

$$\frac{1}{|R|}\sum_{R'\subseteq R}\lambda_{R'}\leqslant 1. \tag{1.7}$$

Again this Bellman function satisfies a certain convexity type inequality, and constructing a function with the same properties allows one to get a proof of the corresponding Carleson imbedding theorem. The function $4(F-\frac{f^2}{1+k})$ is given in [4] (see also [5] for a weighted version).

There are several other problems in Harmonic Analysis where Bellman functions naturally arise. Such problems (including the dyadic Carleson imbedding) are described in [6] (see also [4,5]) and also connections to Stochastic Optimal Control are provided, from which it follows that the corresponding Bellmn functions satisfy certain nonlinear second order PDE, which usually are very difficult to solve.

The exact computation of a Bellman function is a difficult task which is connected with the deeper structure of the corresponding Harmonic Analysis problem. In [6] this is implicitly asked in several cases including the dyadic Carleson imbedding. Thus far except the exact computation of Bellman functions related to the martingale transform, accomplished by Burkholder (see [1,3]) and to the John-Nirenberg inequality, obtained recently by Slavin and Vasyunin [7], no other case of such an exact computation of Bellman functions seems to exist.

The purpose of this paper is to study maximal operators that behave like the dyadic ones (all dyadic ones being special cases) and to exactly compute the corresponding Bellman functions and so in particular compute the Bellman functions defined by (1.5) and (1.6). It will turn out that these functions are the same for all such maximal operators and depend only on the underlying tree structure (see Section 2). The computation of the above Bellman functions will provide refinements of the sharp L^p inequality (1.3), as well as further examples of important Harmonic Analysis problems where the exact form of the corresponding Bellman function is obtained. Our approach will not use the Bellman PDE but will rely on a deeper study of the combinatorial structure of these maximal operators.

Exploiting the remark for replacing M_d by $M_d{}'$ we will work on probability spaces. Hence we will let (X, μ) be a nonatomic probability space and let $\mathcal{T}$ be a family of measurable subsets of X that has a tree-like structure similar to the one in the dyadic case (the precise definition will be given in the next section). Then we can define the maximal operator associated to $\mathcal{T}$ as follows:

$$M_{\mathcal{T}}\phi(x) = \sup\left\{\frac{1}{\mu(I)}\int_I |\phi|\, d\mu : x \in I \in \mathcal{T}\right\} \tag{1.8}$$

for every $\phi \in L^1(X, \mu)$.

The above maximal operator satisfies essentially the same inequalities as M_d, the proof being a trivial adaptation of the proof in the dyadic case. Now we define the corresponding Bellman function as

$$\mathcal{B}_p^{\mathcal{T}}(F,f,L) = \sup\left\{\int_X (\max(M_{\mathcal{T}}\phi, L))^p d\mu : \phi \geqslant 0, \phi \in L^p(X, \mu), \int_X \phi^p\, d\mu = F, \int_X \phi d\mu = f\right\}. \tag{1.9}$$

It is easy to see that when $X = [0,1]^n$, μ is the Lebesgue measure and $\mathcal{T}$ is the family of dyadic cubes contained in X the above function becomes the Bellman function $\mathcal{B}_p$ defined by (1.5) since we can define ϕ on $[0,2]^n \backslash X$ (and set it equal to 0 outside $[0,2]^n$) to make $\sup_{R\,:\, X \subseteq R} \mathrm{Av}_R(\phi) = \mathrm{Av}_{[0,2]^n}(\phi) = L$. Also in the case where $X, \mathcal{T}$ are part of a larger tree-like structure, $\mathcal{B}_p^{\mathcal{T}}$ can be defined in an analogous to (1.5) way.

To state our main result we consider for any $p > 1$ the function

$$H_p(z) = -(p-1)z^p + pz^{p-1} \tag{1.10}$$

defined for $z \in [1, \frac{p}{p-1}]$. It is easy to see that H_p is strictly decreasing on this interval and it maps it onto $[0,1]$. We now let $\omega_p : [0,1] \to [1, \frac{p}{p-1}]$ denote the inverse function H_p^{-1} of H_p. Then we have the following.

Theorem 1. *For any nonatomic probability space (X, μ), any tree-like family $\mathcal{T}$ and any $p > 1$ the corresponding Bellman function is given by*

$$\mathcal{B}_p^{\mathcal{T}}(F,f,L) = \begin{cases} F\omega_p\left(\dfrac{pL^{p-1}f - (p-1)L^p}{F}\right)^p & \text{if } L < \dfrac{p}{p-1}f, \\ L^p + \left(\dfrac{p}{p-1}\right)^p (F - f^p) & \text{if } L \geqslant \dfrac{p}{p-1}f. \end{cases} \tag{1.11}$$

Thus the Bellman functions are the same for any such $X, \mathcal{T}$ and depend only on the underlying tree-like structure. Also note that when $L = f$ the above theorem gives

$$\sup\left\{\int_X (M_{\mathcal{T}}\phi)^p\,d\mu : \phi\geqslant 0, \phi\in L^p(X,\mu), \int_X \phi^p\,d\mu = F, \int_X \phi\,d\mu = f\right\}$$
$$= F\omega_p\left(\frac{f^p}{F}\right)^p. \tag{1.12}$$

which provides a sharp refinement of the best possible $L^p \to L^p$ inequality for the dyadic maximal operators. Actually (1.12) will be an essential step in proving Theorem 1 and will be proved first (see Section 4).

In the particular case $p = 2$, ω_2 can be explicitly computed and we have the following.

Corollary 1. *For any nonatomic probability space* (X,μ) *and any tree-like family* $\mathcal{T}$ *we have*

$$\mathcal{B}_2^{\mathcal{T}}(F,f,L) = \begin{cases} (\sqrt{F} + \sqrt{F + L^2 - 2Lf})^2 & \text{if } L < 2f, \\ L^2 + 4(F - f^2) & \text{if } L \geqslant 2f. \end{cases} \tag{1.13}$$

In particular (*for* $L = f$) *we have* (*with respect to the measure space* (X,μ))

$$||M_{\mathcal{T}}\phi||_2 \leqslant ||\phi||_2 + (||\phi||_2^2 - ||\phi||_1^2)^{1/2} < 2||\phi||_2 \tag{1.14}$$

and this is sharp.

Thus the L^2 norm of the maximal function of any ϕ in L^2 is less than the L^2 norm of ϕ plus the *variance* of ϕ and this gives information on the shape of functions that make the sharp inequality (1.3) for $p = 2$ almost equality (and can be viewed as a rather precise way to express the nonexistence of extremals). Analogous remarks can be made for any $p > 1$.

For the Bellman function associated to the dyadic Carleson imbedding we define $\tilde{\mathcal{B}}_p^{\mathcal{T}}(F,f,k)$ as the supremum of the quantity

$$\sum_{I\in\mathcal{T}} \lambda_I \left(\frac{1}{\mu(I)}\int_I \phi\,d\mu\right)^p, \tag{1.15}$$

where ϕ is nonnegative and in $L^p(X,\mu)$ with $\int_X \phi^p\,d\mu = F$ and $\int_X \phi\,d\mu = f$ and the nonnegative λ_I's satisfy

$$\sum_{J\in\mathcal{T}\,:\,J\subseteq I} \lambda_J \leqslant \mu(I) \quad \text{for every } I\in\mathcal{T} \quad \text{and} \quad \sum_{I\in\mathcal{T}} \lambda_I = k. \tag{1.16}$$

To find this function we show first that

$$\tilde{\mathcal{B}}_p^{\mathcal{T}}(F,f,k) = \sup\left\{ \int_K (M_{\mathcal{T}}\phi)^p \, d\mu : \phi \geqslant 0, \phi \in L^p(X,\mu), \int_X \phi^p \, d\mu = F, \int_X \phi \, d\mu = f \right.$$
$$\left. \text{and } K \subseteq X \text{ is } \mu\text{–measurable with } \mu(K) = k \right\} \tag{1.17}$$

(see Proposition 3 in Section 7). Comparing (1.9) with (1.17) shows the relation between those two Bellman functions when viewed in terms of the maximal operator.

Next, considering $p > 1$ and $0 < k \leqslant 1$ we let for every $U \in [0,1]$ $\omega_{p,k}(U)$ denote the unique (see Lemma 7(i)) solution in the interval $[1, 1 + \frac{k}{p-1}]$ of the equation

$$-(p-1)z^p + (p-1+k)z^{p-1} = U\left[1 + (1-k)\left(\frac{p-1}{z} - p\right)\right]. \tag{1.18}$$

The above Bellman function is then given by the following.

Theorem 2. *We have*

$$\tilde{\mathcal{B}}_p^{\mathcal{T}}(F,f,k) = \left[F\omega_{p,k}\left(\frac{f^p}{F}\right)^p - (1-k)f^p\right]\left[\frac{1-(1-k)\omega_{p,k}\left(\frac{f^p}{F}\right)^{-1}}{k}\right]^p. \tag{1.19}$$

Note that for $k = 1$ we have $\omega_{p,k} = \omega_p$ and so (1.19) becomes (1.12) as expected. Also note that when $X = [0,1]$ and $d\mu = \alpha w \, dx$ where w is any weight (and $\alpha > 0$ to make $\mu(X) = 1$) then the above Theorem (and the independence of $\tilde{\mathcal{B}}_p^{\mathcal{T}}$ from the measure space) allows us to compute the Bellman function associated to the weighted dyadic Carleson imbedding Theorem, defined in [5], as equal to the quantity in (1.19) with $k = M/\mathbf{w}$ (see [5]).

This paper is organized as follows. In Section 2 we define the tree-structures and obtain some basic properties. In Section 3 we collect a number of technical Lemmas needed throughout this paper. In Section 4 we will introduce an effective linearization technique of the maximal operator on an adequate set of functions (which we will call $\mathcal{T}$-good) and use it to prove (1.12). Then in Section 5 we use this result and express each Bellman function as a supremum of a somewhat complicated expression depending on some parameters, and in Section 6 we explicitly compute this supremum and so complete the proof of Theorem 1. In Section 7 we treat the dyadic Carleson imbedding Bellman function, prove first (1.17) and then we apply the above techniques to complete the proof of Theorem 2.

2. Trees and maximal operators

Let (X, μ) be a nonatomic probability space (i.e. $\mu(X) = 1$). Two measurable subsets A, B of X will be called almost disjoint if $\mu(A \cap B) = 0$. Then we give the following.

Definition 1. A set $\mathcal{T}$ of measurable subsets of X will be called a tree if the following conditions are satisfied:

(i) $X \in \mathcal{T}$ and for every $I \in \mathcal{T}$ we have $\mu(I) > 0$.
(ii) For every $I \in \mathcal{T}$ there corresponds a finite or countable subset $\mathcal{C}(I) \subseteq \mathcal{T}$ containing at least two elements such that:
  (a) the elements of $\mathcal{C}(I)$ are pairwise almost disjoint subsets of I,
  (b) $I = \bigcup \mathcal{C}(I)$.
(iii) $\mathcal{T} = \bigcup_{m \geqslant 0} \mathcal{T}_{(m)}$ where $\mathcal{T}_{(0)} = \{X\}$ and $\mathcal{T}_{(m+1)} = \bigcup_{I \in \mathcal{T}_{(m)}} \mathcal{C}(I)$.
(iv) We have $\lim_{m \to \infty} \sup_{I \in \mathcal{T}_{(m)}} \mu(I) = 0$.

Examples. (1) If Q_0 is the unit cube $\mathbb{R}^n$ then the set of all dyadic cubes $Q \subseteq Q_0$ is a tree with $\mathcal{C}(Q)$ being the set of the 2^n subcubes of Q obtained by bisecting its sides. More generally for any integer $m > 1$ the set of all m-adic cubes $Q \subseteq Q_0$ is a tree with $\mathcal{C}(Q)$ being the set of the m^n subcubes of Q obtained by dividing each side of it into m equal parts.

(2) Given the integers $d_1, \ldots, d_n \geqslant 1$ and $m > 1$ we can define a tree $\mathcal{T}$ of parallilepipeds on $X = Q_0$ by setting for each parallilepiped R the family $\mathcal{C}(R)$ to consist of the parallilepipeds formed by dividing the dimensions of R into $m^{d_1}, \ldots, m^{d_n}$ equal parts, respectively. This tree is related to nonisotropic dilations. For example if $n = 2, m = 2, d_1 = 1$ and $d_2 = 2$ we get the set of dyadic parabolic rectangles contained in $[0, 1]^2$.

(3) On $X = [0, 1]$ let $\beta > 0$ and for each $I \subseteq X$ we let $\mathcal{C}(I)$ consist by the two subintervals of I formed by dividing it in ratio β. Then using the relation in (iii) in the above definition we get a tree on X. For $\beta = 1$ we get the dyadic intervals. Actually, it is not hard to see that any tree in a general space X can in a sense be modeled in the space $[0, 1]$ with the Lebesgue measure, but we will not use that.

For any tree $\mathcal{T}$ we define its exceptional set $E = E(\mathcal{T})$ as follows:

$$E(\mathcal{T}) = \bigcup_{I \in \mathcal{T}} \bigcup_{\substack{J_1, J_2 \in \mathcal{C}(I) \\ J_1 \neq J_2}} (J_1 \cap J_2). \tag{2.1}$$

It is clear that $E(\mathcal{T})$ has measure 0.

An easy induction shows that each family $\mathcal{T}_{(m)}$ consists of pairwise almost disjoint sets whose union is X. Moreover if $x \in X \backslash E(\mathcal{T})$ then for each m there exists exactly one $I_m(x)$ in $\mathcal{T}_{(m)}$ containing x. For every $m > 0$ there is a $J \in \mathcal{T}_{(m-1)}$ such that $I_m(x) \in \mathcal{C}(J)$. Since then $x \in J$ we must have $J = I_{m-1}(x)$. Hence the set $\mathcal{A}(x) = \{I \in \mathcal{T} : x \in I\}$ forms a chain $I_0(x) = X \supsetneq I_1(x) \supsetneq \ldots$ with $I_m(x) \in \mathcal{C}(I_{m-1}(x))$ for every

$m > 0$. From this remark it easily follows that if $I, J \in \mathcal{T}$ and $I \cap J \cap (X \backslash E(\mathcal{T}))$ is non-empty then $I \subseteq J$ or $J \subseteq I$. In particular for any $I, J \in \mathcal{T}$ we have either $\mu(I \cap J) = 0$ or one of them is contained in the other. The following gives another property of $\mathcal{T}$ that will be useful later.

Lemma 1. *For every $I \in \mathcal{T}$ and every α such that $0 < \alpha < 1$ there exists a subfamily $\mathcal{F}(I) \subseteq \mathcal{T}$ consisting of pairwise almost disjoint subsets of I such that*

$$\mu\left(\bigcup_{J \in \mathcal{F}(I)} J\right) = \sum_{J \in \mathcal{F}(I)} \mu(J) = (1-\alpha)\mu(I). \tag{2.2}$$

Proof. We will prove the Lemma only for $I = X$ the general case being similar. Using (iv) of Definition 1 we choose m_1 such that $\sup_{I \in \mathcal{T}_{(m_1)}} \mu(I) < \frac{1-\alpha}{2}$ and then choose a subfamily $\mathcal{F}_1 \subseteq \mathcal{G}_1 = \mathcal{T}_{(m_1)}$ with $\sum_{J \in \mathcal{F}_1} \mu(J)$ maximal and such that

$$\sum_{J \in \mathcal{F}_1} \mu(J) \leqslant 1 - \alpha. \tag{2.3}$$

If $\sum_{J \in \mathcal{F}_1} \mu(J) = 1 - \alpha$ we stop. Otherwise since $\sum_{J \in \mathcal{G}_1} \mu(J) = 1$ there exists $I \in \mathcal{G}_1 \backslash \mathcal{F}_1$ and so we must have $\mu(I) + \sum_{J \in \mathcal{F}_1} \mu(J) > 1 - \alpha$ implying

$$0 < \varepsilon_1 = 1 - \alpha - \sum_{J \in \mathcal{F}_1} \mu(J) < \frac{1-\alpha}{2}. \tag{2.4}$$

Next choose $m_2 > m_1$ with $\sup_{I \in \mathcal{T}_{(m_2)}} \mu(I) < \frac{\varepsilon_1}{2}$ let $\mathcal{G}_2$ consist of all I in $\mathcal{T}_{(m_2)}$ that are almost disjoint from $\cup \mathcal{F}_1$ and choose a subfamily $\mathcal{F}_2 \subseteq \mathcal{G}_2$ maximal under

$$\sum_{J \in \mathcal{F}_2} \mu(J) \leqslant \varepsilon_1. \tag{2.5}$$

Again we have

$$0 \leqslant \varepsilon_2 = \varepsilon_1 - \sum_{J \in \mathcal{F}_2} \mu(J) < \frac{\varepsilon_1}{2} < \frac{1-\alpha}{2^2} \tag{2.6}$$

and therefore

$$0 \leqslant 1 - \alpha - \sum_{J \in \mathcal{F}_1 \cup \mathcal{F}_2} \mu(J) < \frac{1-\alpha}{2^2}. \tag{2.7}$$

If equality holds we stop. Otherwise continuing this way we obtain the families $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_s, \ldots$ and it is easy to see that $\mathcal{F} = \bigcup_{s=1}^{\infty} \mathcal{F}_s$ has the desired properties. $\square$

Now given any tree $\mathcal{T}$ we define the maximal operator associated to it as follows:

$$M_{\mathcal{T}}\phi(x) = \sup\left\{\frac{1}{\mu(I)}\int_I |\phi|\,d\mu : x \in I \in \mathcal{T}\right\} \tag{2.8}$$

for every $\phi \in L^1(X, \mu)$.

The above maximal operator satisfies essentially the same inequalities as M_d the proof being a trivial adaptation of the proof in the dyadic case and they are best possible as in the dyadic case.

3. Some technical lemmas

In this section we collect certain technical results that will be used throughout this paper. The first concerns the function H_p and ω_p defined in the introduction.

Lemma 2. *Let $p > 1$ be fixed. Then*

(i) *The function $\omega = \omega_p : [0, 1) \to (1, \frac{p}{p-1}]$ is strictly decreasing and satisfies*

$$\frac{d}{dx}\omega^p(x) = -\frac{1}{p-1}\frac{\omega(x)}{\omega(x)-1}. \tag{3.1}$$

(ii) *The function $G(x) = \omega(x)^p$ is concave in $[0, 1]$. In particular*

$$\omega(x)^p \geqslant x + \left(\frac{p}{p-1}\right)^p (1-x) \tag{3.2}$$

for every $x \in [0, 1]$.

(iii) *The function $U(x) = \frac{\omega(x)^p}{x}$ is strictly decreasing on $(0, 1]$.*

(iv) *For every $x, y \geqslant 0$ we have*

$$(p-1)x^p - px^{p-1}y + y^p \geqslant 0 \tag{3.3}$$

with equality if and only if $x = y$.

Proof. Eq. (3.1) is an easy computation and then (ii) follows since

$$\frac{d^2}{dx^2}\omega^p(x) = \frac{1}{p-1}\frac{\omega'(x)}{(\omega(x)-1)^2} < 0.$$

Now (iii) follows from (ii) since $U(x) = \frac{\omega(x)^p - \omega(0)^p}{x} + \frac{\omega(0)^p}{x}$ is the sum of two decreasing functions and (iv) is well known. $\square$

Lemma 3. *Let $p>1$ and $\tau \in (0,1]$ be fixed. Then for every α with $0<\alpha<1$ the equation*

$$-(z-\alpha)^p + (1-\alpha)^{p-1} z^p = \tau\alpha(1-\alpha)^{p-1} \tag{3.4}$$

has a unique solution $z = z(\alpha,\tau)$ in $[1,+\infty)$ and moreover

$$\lim_{a\to 0^+} z(\alpha,\tau) = \omega_p(\tau). \tag{3.5}$$

Proof. Let $F(\alpha,z) = -(z-\alpha)^p + (1-\alpha)^{p-1}z^p$. Then $F(\alpha,1) = \alpha(1-\alpha)^{p-1} \geqslant \tau\alpha(1-\alpha)^{p-1}$ and $\lim_{z\to+\infty} F(\alpha,z) = -\infty$. On the other hand $\frac{1}{p}\frac{\partial F}{\partial z} = -(z-\alpha)^{p-1} + (1-\alpha)^{p-1}z^{p-1} < 0$ for every $z>1$. Hence (3.4) has a unique solution $z(\alpha,\tau)$ in $[1,+\infty)$. Moreover, it is easy to see that $z(\alpha,\tau) < \frac{\alpha}{1-(1-\alpha)^{(p-1)/p}}$ which is bounded as $\alpha\to 0^+$ and that any limit point z^* of $z(\alpha,\tau)$ as $\alpha \to 0^+$ must satisfy

$$\tau = \lim_{\alpha\to 0^+} \frac{F(\alpha,z^*)}{\alpha} = H_p(z^*)$$

and this proves (3.5). $\square$

For the next lemmas we suppose that $p>1$ is fixed and F, f, L are positive numbers such that $f<L$, $f^p<F$.

Lemma 4. *Consider the equation*

$$C(k) = \frac{(f-kL)^p}{(1-k)^{p-1}} + kL^p = F. \tag{3.6}$$

Then (i) *If $fL^{p-1} \geqslant F$ this equation has a unique solution $k_0 = k_0(F,f,L)$ in the interval $[0,\frac{f}{L}]$ and* (ii) *if $fL^{p-1}<F$ then $C(k)<F$ for any k in $[0,\frac{f}{L}]$.*

Proof. We have

$$C'(k) = L^p\left[(p-1)\left(\frac{f-kL}{L(1-k)}\right)^p - p\left(\frac{f-kL}{L(1-k)}\right)^{p-1} + 1\right] > 0 \tag{3.7}$$

for $0<k<\frac{f}{L}\leqslant 1$ by (3.3), $C(0) = f^p < F$ and $C\left(\frac{f}{L}\right) = fL^{p-1}$. $\square$

Lemma 5. *Consider the equation*

$$D(k) = (p-1)\left(\frac{L(1-k)}{f-kL}\right)^p - p\left(\frac{L(1-k)}{f-kL}\right)^{p-1} = \frac{(p-1)L^p - pfL^{p-1}}{F}. \tag{3.8}$$

Then (i) *If* $L \leqslant \frac{p}{p-1} f$ *this equation has a unique solution* $k_1 = k_1(F, f, L)$ *in the interval* $[0, \frac{f}{L}]$ *and* (ii) *if* $L > \frac{p}{p-1} f$ *then* $D(k) > \frac{(p-1)L^p - pfL^{p-1}}{F}$ *for any* k *in* $[0, \frac{f}{L}]$.

Proof. Setting $\xi(k) = \frac{L(1-k)}{f-kL}$, ξ is strictly increasing and maps $[0, \frac{f}{L})$ onto $[\frac{L}{f}, +\infty)$. Also $D(k) = g(\xi(k))$ where $g(\xi) = (p-1)\xi^p - p\xi^{p-1}$ has $g'(\xi) = p(p-1)\xi^{p-2}(\xi - 1) > 0$ for $\xi > 1$. Since

$$g\left(\frac{L}{f}\right) \leqslant \frac{(p-1)L^p - pfL^{p-1}}{F} \tag{3.9}$$

if and only if $(F - f^p)((p-1)L^p - pfL^{p-1}) \leqslant 0$ and since $F > f^p$ the proof is complete. $\square$

Lemma 6. *Suppose that* $L \leqslant \frac{p}{p-1} f$ *and* $fL^{p-1} \geqslant F$. *Then*

$$k_1(F, f, L) < k_0(F, f, L) \tag{3.10}$$

Proof. By (3.7) it suffices to prove that $C(k_1) < C(k_0) = F$. Writing

$$\xi = \frac{L(1-k_1)}{f - k_1 L}, \quad \varDelta = \frac{(p-1)L^p - pfL^{p-1}}{F}$$

the inequality

$$C(k_1) = L^p \left[\frac{L - f}{L(\xi - 1)\xi^{p-1}} + \frac{f\xi - L}{L(\xi - 1)} \right] < F$$

is equivalent to $L^p - L^{p-1} f + L^{p-1}(f\xi^p - L\xi^{p-1}) < F(\xi^p - \xi^{p-1})$ and since $\xi^p = \frac{p}{p-1}\xi^{p-1} + \frac{\varDelta}{p-1}$ to

$$fL^{p-1}\frac{\varDelta + 1}{p-1} < \frac{F(\varDelta + 1)}{p-1}\xi^{p-1}.$$

But $F(\varDelta + 1) > f^p - pfL^{p-1} + (p-1)L^p > 0$ by (3.3) and so it suffices to show that $\xi^{p-1} > \frac{fL^{p-1}}{F} (\geqslant 1)$ and for this it suffices to show that $g((\frac{f}{F})^{1/(p-1)} L) < g(\xi) = \varDelta$ which can be written as $(p-1)(\frac{f}{F})^{p/(p-1)} L^p - p\frac{fL^{p-1}}{F} < (p-1)\frac{L^p}{F} - p\frac{fL^{p-1}}{F}$ and this follows easily since $f^p < F$. $\square$

Now for the next lemma we also fix k with $0 < k < 1$ and we consider the functions

$$h_k(B) = \frac{(f - B)^p}{(1-k)^{p-1}} + \frac{B^p}{k^{p-1}} \tag{3.11}$$

defined for $0 \leqslant B \leqslant f$ and

$$R_k(B) = \left(F - \frac{(f-B)^p}{(1-k)^{p-1}}\right)\omega_p\left(\frac{B^p}{k^{p-1}(F - \frac{(f-B)^p}{(1-k)^{p-1}})}\right)^p \tag{3.12}$$

defined for all $B \in [0,f]$ such that $h(B) \leqslant F$.

Noting that h_k has an absolute minimum at $B = kf$ with $h_k(kf) = f^p < F$ and that it is monotone on each of the intervals $(0,kf)$ and (kf,f) we conclude that either $h_k(f) > F$ i.e. $f^p > k^{p-1}F$ in which case the equation $h_k(B) = F$ has a unique solution in (kf,f) and this is denoted by $\rho_1(F,f,k)$ or $h_k(f) \leqslant F$ in which case we set $\rho_1(F,f,k) = f$. Also either $h_k(0) > F$ i.e. $f^p > (1-k)^{p-1}F$ in which case the equation $h_k(B) = F$ has a unique solution in $(0,kf)$ and this is denoted by $\rho_0(F,f,k)$ or $h_k(0) \leqslant F$ in which case we set $\rho_0(F,f,k) = 0$. Clearly in all cases the domain of definition of R_k is the interval $[\rho_0(F,f,k), \rho_1(F,f,k)]$. We now have the following.

Lemma 7. (i) *For every $U \in [0,1]$ the equation*

$$\sigma(z) = -(p-1)z^p + (p-1+k)z^{p-1} - U\left[1 + (1-k)\left(\frac{p-1}{z} - p\right)\right] = 0 \tag{3.13}$$

has a unique solution in the interval $[1, 1+\frac{k}{p-1}]$ which is denoted by $\omega_{p,k}(U)$.

(ii) *The function R_k defined on $[\rho_0(F,f,k), \rho_1(F,f,k)]$ assumes its absolute maximum at the unique interior point $B_0 \in (kf, \min(\frac{pk}{p-1+k}f, \rho_1(F,f,k)))$ such that*

$$\frac{f(1-k)}{f - B_0} = \omega_{p,k}\left(\frac{f^p}{F}\right). \tag{3.14}$$

Moreover

$$R_k(B_0) = \max R_k = \left[F\omega_{p,k}\left(\frac{f^p}{F}\right)^p - (1-k)f^p\right]\left[\frac{1-(1-k)\omega_{p,k}(\frac{f^p}{F})^{-1}}{k}\right]^p. \tag{3.15}$$

Proof. (i) We have $\sigma(1) = k(1-U) \geqslant 0$, $\sigma(1+\frac{k}{p-1}) = -\frac{pk^2}{p+k-1}U \leqslant 0$ and

$$\sigma'(z) = -\frac{p-1}{z^2}[z^p(pz - (p-1+k)) - (1-k)U] < 0$$

whenever $z > 1$.

(ii) Setting $Z = \frac{B^p}{k^{p-1}(F - \frac{(f-B)^p}{(1-k)^{p-1}})}$ and using Lemma 2(i) and the equality $Z = -(p-1)\omega_p(Z)^p + p\omega_p(Z)^{p-1}$ it is easy to compute

$$R_k{}'(B) = \frac{p}{p-1}\frac{\omega_p(Z)}{\omega_p(Z)-1}\left[\omega_p(Z)^{p-1}\left(\frac{f-B}{1-k}\right)^{p-1} - \left(\frac{B}{k}\right)^{p-1}\right]. \tag{3.16}$$

Hence using that $1 \leqslant \omega_p(Z) \leqslant \frac{p}{p-1}$ we conclude that $R_k{}'(B) > 0$ whenever $\frac{f-B}{1-k} > \frac{B}{k}$ that is whenever $B < kf$ and that $R_k(B) < 0$ whenever $\frac{p}{p-1}\frac{f-B}{1-k} < \frac{B}{k}$ that is whenever $B > \frac{pk}{p-1+k}f$. Moreover we have $\lim_{B \to \rho_1(F,f,k)} R_k{}'(B) < 0$ (it is $-\infty$ if $\rho_1 < f$). Hence $R_k{}'$ has at least one zero B_0 in the open interval $(kf, \min(\frac{pk}{p-1+k}f, \rho_1(F,f,k)))$ such that $R_k(B_0)$ is the absolute maximum of R_k. But from (3.16) it follows that $R_k{}'(B) = 0$ if and only if

$$\omega_p(Z) = \frac{B}{k}\frac{1-k}{f-B} \tag{3.17}$$

hence $Z = H_p(\frac{B}{k}\frac{1-k}{f-B})$ and setting $z = \frac{f(1-k)}{f-B}$ this is equivalent to (3.13) with $U = \frac{f^p}{F}$. This completes the proof, since (3.15) now follows from an easy computation using (3.17). $\square$

4. Linearization

In this section we describe an effective linearization for the operator $M_{\mathcal{T}}$ valid for certain good functions ϕ. This will be important for proving Theorem 1.

Let $\phi \in L^1(X, \mu)$ be a nonnegative function and for any $I \in \mathcal{T}$ let

$$\mathrm{Av}_I(\phi) = \frac{1}{\mu(I)} \int_I |\phi| d\mu \tag{4.1}$$

We will say that ϕ is $\mathcal{T}$-*good* if the set

$$\Lambda_\phi = \{x \in X \backslash E(\mathcal{T}) : M_{\mathcal{T}}\phi(x) > \mathrm{Av}_I(\phi) \text{ for all } I \in \mathcal{T} \text{ such that } x \in I\} \tag{4.2}$$

has μ-measure zero.

Examples. (1) If $m \geqslant 0$ and $\lambda_I \geqslant 0$ for each $I \in \mathcal{T}_{(m)}$ are given then the function $\phi = \sum_{I \in \mathcal{T}_{(m)}} \lambda_I \chi_I$ (where χ_I denotes the characteristic function of I) is $\mathcal{T}$-good since $\mathrm{Av}_J(\phi) = \mathrm{Av}_P(\phi)$ whenever $J \in \mathcal{C}(P)$, $P \in \mathcal{T}_{(s)}$ for $s > m$ and so $\Lambda_\phi = \emptyset$. We call such a ϕ a $\mathcal{T}$-step function.

(2) If $\mathcal{T}$ is a tree consisting of subintervals of $X = [0,1]$ (with μ being the Lebesgue measure) then it is easy to construct (using Cantor like sets) a set $H \subseteq [0,1]$ of the first category such that $|H \cap I| > 0$ and $|(X \backslash H) \cap I| > 0$ for every nonempty interval $I \subseteq [0,1]$. Then considering the function $\phi = \chi_H$ we have whenever $x \in H$ and $x \in I \in \mathcal{T}$

$$|I|\phi(x) = |I| > |H \cap I| = \int_I \phi \tag{4.3}$$

hence $H \backslash E(\mathcal{T})$ is contained in Λ_ϕ and therefore ϕ is *not* $\mathcal{T}$-good.

Suppose now that ϕ is $\mathcal{T}$-good. Then for any $x \in X \backslash (E(\mathcal{T}) \cup \Lambda_\phi)$ (i.e. for μ-almost every x in X) we define $I_\phi(x)$ to be the *largest* element in the nonempty set $\{I \in \mathcal{T}\colon x \in I \text{ and } M_{\mathcal{T}}\, \phi(x) = \mathrm{Av}_I(\phi)\}$.

Also given any $I \in \mathcal{T}$ let

$$A(\phi, I) = \{x \in X \backslash (E(\mathcal{T}) \cup \Lambda_\phi) : I_\phi(x) = I\} \subseteq I \tag{4.4}$$

and let

$$\mathcal{S}_\phi = \{I \in \mathcal{T} : \mu(A(\phi, I)) > 0\} \cup \{X\}. \tag{4.5}$$

It is now clear that

$$M_{\mathcal{T}}\, \phi = \sum_{I \in \mathcal{S}_\phi} \mathrm{Av}_I(\phi) \chi_{A(\phi, I)} \text{ almost everywhere.} \tag{4.6}$$

We also define the correspondence $I \to I^*$ with respect to $\mathcal{S}_\phi$ as follows: I^* is the smallest element of $\{J \in \mathcal{S}_\phi : I \subsetneq J\}$. This is defined for every I in $\mathcal{S}_\phi$ except X.

The sets $\mathcal{S}_\phi$ and $A(\phi, I)$ will be important in analysing $M_{\mathcal{T}}\, \phi$. It is clear that the $A(\phi, I)$'s for $I \in \mathcal{S}_\phi$ are pairwise disjoint and since obviously $\mu(\bigcup_{J \notin \mathcal{S}_\phi} A(\phi, J)) = 0$, their union has full measure.

Next, we will give some further properties of these sets which will be important in all that follows. We will say that two measurable subsets A, B of X are almost equal and write $A \approx B$ if $\mu(A \backslash B) = \mu(B \backslash A) = 0$.

Lemma 8. (i) *If $I, J \in \mathcal{S}_\phi$ then either $A(\phi, J) \cap I = \emptyset$ or $J \subseteq I$.*

(ii) *If $I \in \mathcal{S}_\phi$ then there exists $J \in \mathcal{C}(I)$ such that $J \notin \mathcal{S}_\phi$.*

(iii) *For every $I \in \mathcal{S}_\phi$ we have $I \approx \bigcup_{\mathcal{S}_\phi \ni J \subseteq I} A(\phi, J)$.*

(iv) *For every $I \in \mathcal{S}_\phi$ we have $A(\phi, I) \approx I \backslash \bigcup_{J \in \mathcal{S}_\phi\colon\ J^* = I} J$ and so*

$$\mu(A(\phi, I)) = \mu(I) - \sum_{J \in \mathcal{S}\colon\ J^* = I} \mu(J). \tag{4.7}$$

Proof. (i) Supposing that $x \in A(\phi, J) \cap I$ we have $x \in I \cap J \neq \emptyset$ and since $x \notin E(\mathcal{T})$ we have either $I \subseteq J$ or $J \subseteq I$. Suppose now that $I \subsetneq J$. Then also $\mathrm{Av}_J(\phi) = M_{\mathcal{T}} \phi(x) \geqslant \mathrm{Av}_I(\phi)$ and so I cannot be an $I_\phi(z)$ for any $z \in I$. Therefore $A(\phi, I) = \emptyset$ contradicting the assumption $I \in \mathcal{S}_\phi$. Hence we must have $J \subseteq I$.

(ii) Since

$$\mathrm{Av}_I(\phi) = \frac{\sum_{J \in \mathcal{C}(I)} \mu(J) \mathrm{Av}_J(\phi)}{\sum_{J \in \mathcal{C}(I)} \mu(J)} \tag{4.8}$$

there exists $J \in \mathcal{C}(I)$ such that $\mathrm{Av}_J(\phi) \leqslant \mathrm{Av}_I(\phi)$ and since $J \subsetneq I$ we have $J \notin \mathcal{S}_\phi$.

(iii) Suppose that $x \in I \backslash (E(\mathcal{T}) \cup \Lambda_\phi)$ and let $J = I_\phi(x)$. Then $x \in A(\phi, J) \cap I \neq \emptyset$ hence by (i) we have $J \subseteq I$. Now the proof is completed observing that $\bigcup_{J \notin \mathcal{S}_\phi} A(\phi, J)$ has measure zero.

(iv) Using (iii) we can write $I \setminus \bigcup_{J \in \mathcal{S}_\phi : J^* = I} J \approx \bigcup \{A(\phi, P) : P \in \mathcal{S}_\phi, P \subseteq I$ and $P \nsubseteq J$ for any $J \in \mathcal{S}_\phi$ with $J^* = I\} = A(\phi, I)$. $\square$

From the above lemma we have in particular

$$\mathrm{Av}_I(\phi) = \frac{1}{\mu(I)} \sum_{J \in \mathcal{S}_\phi : J \subseteq I} \int_{A(\phi,J)} \phi d\mu. \tag{4.9}$$

Now fix $p > 1$. Setting

$$x_I = a_I^{-1+\frac{1}{p}} \int_{A(\phi,I)} \phi \, d\mu \tag{4.10}$$

for every $I \in \mathcal{S}_\phi$ where $a_I = \mu(A(\phi, I))$ (in the case where $\mu(A(\phi, X)) = 0$ we set $x_X = 0$) the above lemma and Hölder's inequality imply that

$$M_{\mathcal{T}}\, \phi = \sum_{I \in \mathcal{S}_\phi} \left(\frac{1}{\mu(I)} \sum_{J \in \mathcal{S}_\phi, J \subseteq I} a_J^{1/q} x_J \right) \chi_{A(\phi,I)} \tag{4.11}$$

almost everywhere, where $q = \frac{p}{p-1}$ is the dual exponent of p,

$$\int_X \phi^p \, d\mu = \sum_{I \in \mathcal{S}_\phi} \int_{A(\phi,I)} \phi^p \, d\mu \geqslant \sum_{I \in \mathcal{S}_\phi} x_I^p \tag{4.12}$$

and

$$\int_X \phi \, d\mu = \sum_{I \in \mathcal{S}_\phi} \int_{A(\phi,I)} \phi \, d\mu = \sum_{I \in \mathcal{S}_\phi} a_I^{1/q} x_I. \tag{4.13}$$

Now we have

$$\int_X (M_{\mathcal{T}}\phi)^p \, d\mu = \sum_{I \in \mathcal{S}_\phi} \left(a_I^{1/p} \frac{1}{\mu(I)} \sum_{J \in \mathcal{S}_\phi, J \subseteq I} a_J^{1/q} x_J \right)^p \tag{4.14}$$

and our first step is to prove the following.

Lemma 9. *If the nonnegative function $\phi \in L^p(X, \mu)$ is $\mathcal{T}$-good $\int_X \phi^p \, d\mu = F$ and $\int_X \phi \, d\mu = f$ then we have*

$$\int_X (M_{\mathcal{T}}\phi)^p \, d\mu \leqslant F \omega_p \left(\frac{f^p}{F} \right)^p. \tag{4.15}$$

Proof. We let $\mathcal{S} = \mathcal{S}_\phi$, $a_I = \mu(A(\phi, I))$,

$$\rho_I = \frac{a_I}{\mu(I)} \in (0,1) \tag{4.16}$$

(except possibly for $I = X$) and

$$y_I = \mathrm{Av}_I(\phi) = \frac{1}{\mu(I)} \sum_{J \in \mathcal{S}\,:\, J \subseteq I} a_J^{1/q} x_J \tag{4.17}$$

for every $I \in \mathcal{S}$. It is easy to see that

$$x_I = a_I^{-1/q}\Big(y_I \mu(I) - \sum_{J \in \mathcal{S}:\, J^* = I} y_J \mu(J)\Big) \tag{4.18}$$

and so using (4.12), (4.7) and Hölder's inequality in the following form:

$$\frac{(\lambda_1 + \cdots + \lambda_m)^p}{(\sigma_1 + \cdots + \sigma_m)^{p-1}} \leqslant \frac{\lambda_1^p}{\sigma_1^{p-1}} + \cdots + \frac{\lambda_m^p}{\sigma_m^{p-1}} \tag{4.19}$$

we conclude that for any $\beta \geqslant 0$ we have

$$\begin{aligned}
F &\geqslant \sum_{I \in \mathcal{S}} \frac{(y_I \mu(I) - \sum_{J^*=I} y_J \mu(J))^p}{(\mu(I) - \sum_{J^*=I} \mu(J))^{p-1}} \\
&\geqslant \sum_{I \in \mathcal{S}} \left(\frac{(y_I \mu(I))^p}{(\tau_I \mu(I))^{p-1}} - \sum_{J^*=I} \frac{(y_J \mu(J))^p}{((\beta+1)\mu(J))^{p-1}} \right) \\
&= \sum_{I \in \mathcal{S}} \frac{(y_I \mu(I))^p}{(\tau_I \mu(I))^{p-1}} - \sum_{\substack{I \in \mathcal{S} \\ I \neq X}} \frac{(y_I \mu(I))^p}{((\beta+1)\mu(I))^{p-1}} \\
&= \frac{y_X^p}{\tau_X^{p-1}} + \sum_{\substack{I \in \mathcal{S} \\ I \neq X}} \frac{1}{\rho_I} \left(\frac{1}{\tau_I^{p-1}} - \frac{1}{(\beta+1)^{p-1}} \right) a_I y_I^p
\end{aligned} \tag{4.20}$$

provided that the $\tau_I > 0$ satisfy $\tau_I \mu(I) - (\beta+1) \sum_{J^*=I} \mu(J) = \mu(I) - \sum_{J^*=I} \mu(J)$ which give

$$\tau_I = \beta + 1 - \beta \rho_I > 0 \tag{4.21}$$

and so

$$F \geqslant \frac{y_X^p}{\tau_X^{p-1}} + \sum_{\substack{I \in \mathcal{S} \\ I \neq X}} \frac{1}{\rho_I} \left(\frac{1}{(\beta + 1 - \beta\rho_I)^{p-1}} - \frac{1}{(\beta+1)^{p-1}} \right) a_I y_I^p. \tag{4.22}$$

Now we note that (by the mean value theorem)

$$\frac{1}{(\beta+1-\beta x)^{p-1}}-\frac{1}{(\beta+1)^{p-1}}\geqslant\frac{(p-1)\beta x}{(\beta+1)^{p}} \tag{4.23}$$

for all $x\in[0,1]$ and we use (4.13), (4.14) to get

$$\begin{aligned} F &\geqslant \frac{y_X^p}{\tau_X^{p-1}}+\frac{(p-1)\beta}{(\beta+1)^p}\sum_{\substack{I\in\mathcal{S}\\ I\neq X}} a_I y_I^p \\ &= \left(\frac{1}{(\beta+1-\beta\rho_X)^{p-1}}-\frac{(p-1)\beta\rho_X}{(\beta+1)^p}\right)f^p+\frac{(p-1)\beta}{(\beta+1)^p}\int_X (M_{\mathcal{T}}\phi)^p\,d\mu \end{aligned}$$

and next we use (4.23) with $x=\rho_X$ to get

$$F\geqslant\frac{1}{(\beta+1)^{p-1}}f^p+\frac{(p-1)\beta}{(\beta+1)^p}\int_X (M_{\mathcal{T}}\phi)^p\,d\mu \tag{4.24}$$

which gives that for any $\beta\geqslant 0$ we have

$$\int_X (M_{\mathcal{T}}\phi)^p\,d\mu\leqslant\left(1+\frac{1}{\beta}\right)\frac{(\beta+1)^{p-1}F-f^p}{p-1}. \tag{4.25}$$

Considering now the right-hand side of (4.25) as a function of β it is easy to see that it is minimized for β equal to the unique root of the equation

$$H_p(\beta+1)=-(p-1)(\beta+1)^p+p(\beta+1)^{p-1}=\frac{f^p}{F}\leqslant 1 \tag{4.26}$$

and so taking $\beta=\omega_p(\frac{f^p}{F})-1\geqslant 0$ in (4.25) and using (4.26) completes the proof of (4.15). $\square$

Remark. In case $a_X=0$ the above proof has to be modified. The inequality for $I=X$ in the first line of (4.20) will now be $y_X^p-\sum_{J^*=I}y_J^p\mu(J)\leqslant 0$ (which holds since $x_X=a_X=0$ imply $\sum_{J^*=I}y_J\mu(J)=y_X$ and $\sum_{J^*=I}\mu(J)=1$) divided by $(\beta+1)^{p-1}$. However we will not need this modification since (as Lemma 8(ii) implies) we have $a_X>0$ whenever ϕ is a $\mathcal{T}$-step function, and this will be enough as we will see in the proof of the following Theorem.

Next we show that (4.15) holds for general ϕ and that it is actually best possible. This is the content of the following.

Theorem 3. *For any $p > 1$ we have*

$$\sup\left\{ \int_X (M_{\mathcal{T}}\phi)^p \, d\mu \colon \phi \geqslant 0, \phi \in L^p(X,\mu), \int_X \phi^p \, d\mu = F, \int_X \phi \, d\mu = f \right\}$$
$$= F\omega_p\left(\frac{f^p}{F}\right)^p. \tag{4.27}$$

Proof. For the general nonnegative $\phi \in L^p(X,\mu)$ one can consider the sequence (ϕ_m) where $\phi_m = \sum_{I \in \mathcal{T}_{(m)}} \mathrm{Av}_I(\phi)\chi_I$ and set

$$\Phi_m = \sum_{I \in \mathcal{T}_{(m)}} \max\{\mathrm{Av}_J(\phi) : I \subseteq J \in \mathcal{T}\}\chi_I = M_{\mathcal{T}}\phi_m$$

since $\mathrm{Av}_J(\phi) = \mathrm{Av}_J(\phi_m)$ whenever $I \subseteq J \in \mathcal{T}$.

Then it is easy to see that

$$\int_X \phi_m \, d\mu = \int_X \phi \, d\mu = f, \quad F_m = \int_X \phi_m^p \, d\mu \leqslant \int_X \phi^p \, d\mu = F \tag{4.28}$$

for all m and that Φ_m converges monotonically almost everywhere to $M_{\mathcal{T}}\phi$. Since as we have seen each ϕ_m is $\mathcal{T}$-good, (4.15) combined with Lemma 2(iii) gives

$$\int_X \Phi_m^p \, d\mu \leqslant F_m \omega_p\left(\frac{f^p}{F_m}\right)^p \leqslant F\omega_p\left(\frac{f^p}{F}\right)^p \tag{4.29}$$

and so letting $m \to \infty$ we get (4.15) for the general ϕ.

Now to complete the proof of the theorem we choose α with $0 < \alpha < 1$ and using Lemma 1, for every $I \in \mathcal{T}$ we choose a family $\mathcal{F}(I) \subseteq \mathcal{T}$ of pairwise almost disjoint subsets of I such that

$$\sum_{J \in \mathcal{F}(I)} \mu(J) = (1-\alpha)\mu(I). \tag{4.30}$$

Then we define $\mathcal{S}$ to be the smallest subset of $\mathcal{T}$ such that $X \in \mathcal{S}$ and for every $I \in \mathcal{S}$, $\mathcal{F}(I) \subseteq \mathcal{S}$. It is clear that defining the correspondence $I \to I^*$ with respect to this $\mathcal{S}$ we have $J^* = I \in \mathcal{S}$ if and only if $J \in \mathcal{F}(I)$ and so writing

$$A_I = I \Big\backslash \bigcup_{J \in \mathcal{S} \,:\, J^* = I} J \tag{4.31}$$

we have $a_I = \mu(A_I) = \mu(I) - \sum_{J \in \mathcal{S} \,:\, J^* = I} \mu(J) = \alpha\mu(I)$ for every $I \in \mathcal{S}$.

Also it is easy to see that

$$\mathcal{S} = \bigcup_{m \geqslant 0} \mathcal{S}_{(m)} \text{ where } \mathcal{S}_{(0)} = \{X\} \text{ and } \mathcal{S}_{(m+1)} = \bigcup_{I \in \mathcal{S}_{(m)}} \mathcal{F}(I). \tag{4.32}$$

We define the rank $r(I)$ of any $I \in \mathcal{S}$ to be the unique integer m such that $I \in \mathcal{S}_{(m)}$ and for $\lambda, \gamma > 0$ to be chosen later we define the x_I's by setting

$$x_I = \lambda \gamma^{r(I)} \mu(I)^{1/p} \tag{4.33}$$

for every $I \in \mathcal{S}$. For every $I \in \mathcal{S}$ and every $m \geqslant 0$ we write

$$b_m(I) = \sum_{\substack{\mathcal{S} \ni J \subseteq I \\ r(J) = r(I) + m}} \mu(J) \tag{4.34}$$

and observing that

$$b_{m+1}(I) = \sum_{\substack{\mathcal{S} \ni J \subseteq I \\ r(J) = r(I) + m}} \sum_{L \in \mathcal{F}(J)} \mu(L) = (1 - \alpha) b_m(I) \tag{4.35}$$

we get

$$b_m(I) = (1 - \alpha)^m \mu(I). \tag{4.36}$$

Hence

$$\begin{aligned} \sum_{I \in \mathcal{S}} x_I^p &= \lambda^p \sum_{m \geqslant 0} \sum_{I \in \mathcal{S}_{(m)}} \gamma^{mp} \mu(J) = \lambda^p \sum_{m \geqslant 0} \gamma^{mp} b_m(X) \\ &= \lambda^p \sum_{m \geqslant 0} [\gamma^p (1 - \alpha)]^m = \frac{\lambda^p}{1 - \gamma^p (1 - \alpha)}, \end{aligned} \tag{4.37}$$

where we assume that $\gamma^p (1 - \alpha) < 1$. On the other hand

$$\begin{aligned} a_I^{1/p} y_I &= \frac{1}{\mu(I)} \sum_{J \in \mathcal{S} : J \subseteq I} a_I^{1/p} a_J^{1/q} x_J \\ &= \frac{\lambda}{\mu(I)} \sum_{\mathcal{S} \ni J \subseteq I} (\alpha \mu(I))^{1/p} (\alpha \mu(J))^{1/q} \gamma^{r(J)} \mu(J)^{1/p} \\ &= \lambda \alpha \mu(I)^{-1+1/p} \sum_{m \geqslant 0} \gamma^{m + r(I)} \sum_{\substack{\mathcal{S} \ni J \subseteq I \\ r(J) = r(I) + m}} \mu(J) \\ &= \lambda \alpha \gamma^{r(I)} \mu(I)^{-1+1/p} \sum_{m \geqslant 0} \gamma^m (1 - \alpha)^m \mu(I) = \frac{\alpha}{1 - \gamma(1 - \alpha)} x_I, \end{aligned}$$

where the y_I's are defined by the first equality above and so

$$\sum_{I \in \mathcal{S}} a_I y_I^p = \left(\frac{\alpha}{1 - \gamma(1 - \alpha)} \right)^p \sum_{I \in \mathcal{S}} x_I^p. \tag{4.38}$$

whereas

$$y_X = \frac{\lambda \alpha^{1/q}}{1-\gamma(1-\alpha)}. \tag{4.39}$$

We choose $\lambda = f\alpha^{-1/q}(1-\gamma(1-\alpha))$ to make $y_X = f$ and $\gamma = \frac{z-\alpha}{z(1-\alpha)}$ where $z = z(\alpha, \frac{f^p}{F}) > 1$ is the unique root of $\frac{-(z-\alpha)^p + (1-\alpha)^{p-1} z^p}{\alpha(1-\alpha)^{p-1}} = \frac{f^p}{F}$ furnished by Lemma 3. Obviously

$$\gamma^p(1-\alpha) = \frac{(z-\alpha)^p}{z^p(1-\alpha)^{p-1}} = 1 - \frac{f^p}{F}\frac{\alpha}{z^p} < 1$$

and so we have for these choices of λ, γ

$$\sum_{I \in \mathcal{S}} x_I^p = F, y_X = f \quad \text{and} \quad \sum_{I \in \mathcal{S}} a_I y_I^p = z\left(\alpha, \frac{f^p}{F}\right)^p F. \tag{4.40}$$

Next, we consider the function

$$\phi_\alpha = \sum_{I \in \mathcal{S}} \frac{x_I}{a_I^{1/p}} \chi_{A_I}. \tag{4.41}$$

Then it is easy to see that $\int_X \phi_\alpha^p \, d\mu = F, \int_X \phi_\alpha \, d\mu = f$ and since

$$M_{\mathcal{T}}\phi_\alpha \geqslant \sum_{I \in \mathcal{S}} \mathrm{Av}_I(\phi_\alpha)\chi_{A_I} = \sum_{I \in \mathcal{S}} \left(\frac{1}{\mu(I)} \sum_{J \in \mathcal{S}\,:\, J \subseteq I} a_J^{1/q} x_J \right) \chi_{A_I} \tag{4.42}$$

we have

$$\int_X (M_{\mathcal{T}}\phi_\alpha)^p \, d\mu \geqslant \sum_{I \in \mathcal{S}} a_I y_I^p = z\left(\alpha, \frac{f^p}{F}\right)^p F. \tag{4.43}$$

Now we let $\alpha \to 0^+$ and using Lemma 3 completes the proof. $\square$

Using similar arguments we can also prove the following

Theorem 4. *Let $p > 1$ and suppose α is such that $0 < \alpha < 1$. If ϕ is $\mathcal{T}$-good and such that $a_I = \mu(A(\phi, I)) \geqslant \alpha\mu(I)$ for every $I \in \mathcal{S} = \mathcal{S}_\phi$ then*

$$\int_X (M_{\mathcal{T}}\phi)^p \, d\mu \leqslant \left(\frac{\alpha}{1-(1-\alpha)^{(p-1)/p}} \right)^p \int_X \phi^p \, d\mu < \left(\frac{p}{p-1} \right)^p \int_X \phi^p \, d\mu \tag{4.44}$$

and this is best possible.

Proof. To prove (4.44) we let $F = \int_X \phi^p \, d\mu$ and arguing as in the proof of Lemma 9 we arrive at (4.22). Next noting that for any fixed $\beta \geqslant 0$ the function

$$g_\beta(x) = \frac{1}{x}\left(\frac{1}{(\beta+1-\beta x)^{p-1}} - \frac{1}{(\beta+1)^{p-1}}\right) \tag{4.45}$$

is increasing in $(0,1)$ (since $h_\beta(x) = (\beta+1-\beta x)^{-(p-1)}$ is convex) we get using that each $\rho_I \geqslant \alpha$ that

$$F \geqslant y_X^p \left(\frac{1}{(\beta+1-\beta\rho_X)^{p-1}} - \rho_X g_\beta(\alpha)\right) + g_\beta(\alpha) \sum_{I \in \mathcal{S}} a_I y_I^p. \tag{4.46}$$

But $\frac{1}{(\beta+1-\beta\rho_X)^{p-1}} - \rho_X g_\beta(\alpha) > 0$ since $\frac{h_\beta(\rho_X)-h_\beta(0)}{\rho_X} \geqslant \frac{h_\beta(\alpha)-h_\beta(0)}{\alpha}$ and therefore

$$\int_X (M_{\mathcal{T}}\phi)^p \, d\mu \leqslant \frac{1}{g_\beta(\alpha)} F \tag{4.47}$$

and so choong β with $\frac{\beta}{\beta+1}\alpha = 1-(1-\alpha)^{1/p} < \alpha$ completes the proof of (4.44).

To show that (4.44) is best possible we define $\mathcal{S}$ the A_I's and the x_I's as in the proof of Theorem 3, with $\lambda = 1$ and $\gamma > 0$ such that $\gamma^p(1-\alpha) < 1$. Defining ϕ_α and arguing as before we get

$$\int_X (M_{\mathcal{T}}\phi_\alpha)^p \, d\mu \geqslant \left(\frac{\alpha}{1-\gamma(1-\alpha)}\right)^p \int_X \phi_\alpha^p \, d\mu. \tag{4.48}$$

Letting now $\gamma \to (1-\alpha)^{-1/p}$ completes the proof. □

The above theorem has the following application in the case where $X = [0,1]$, μ is the Lebesgue measure and $\mathcal{T} = \mathcal{D}$ is the tree of all dyadic subintervals of $[0,1]$ and therefore $\mathcal{D}_{(m)}$ is the set of all dyadic subintervals of $[0,1]$ that have length equal to 2^{-m}.

Proposition 1. *Let* $p > 1$. *Then*:

(i) *If* $m > 0$ *and* $\phi = \sum_{I \in \mathcal{D}_{(m)}} \lambda_I \chi_I$ *then*

$$||M_{\mathcal{D}}\phi||_p \leqslant \frac{1}{2^m - 2^{m/p}(2^m-1)^{(p-1)/p}} ||\phi||_p. \tag{4.49}$$

(ii) *If* $\phi \in L^p$ *and is nonnegative, decreasing and convex on* $(0,1)$ *then*

$$||M_{\mathcal{D}}\phi||_p \leqslant \frac{1}{2-2^{1/p}} ||\phi||_p. \tag{4.50}$$

Proof. Both follow from (4.44). For (4.49) it suffices to observe that ϕ is $\mathcal{D}$-good and for any $J \in \mathcal{S}_\phi$ we choose $I \in \mathcal{D}_{(m)}$ such that $|A(\phi, J) \cap I| > 0$ and note that we must have $I \subseteq A(\phi, J)$ and so $\frac{|A(\phi,J)|}{|J|} \geqslant 2^{-m}$.

For (4.50) we use (4.44) for $\alpha = \frac{1}{2}$ noting that for any $J \in \mathcal{D}$ if J_+ is the right half of J, the assumptions on ϕ imply

$$\mathrm{Av}_J(\phi) \geqslant \sup_{x \in J_+} \phi(x) \tag{4.51}$$

and therefore ϕ is $\mathcal{D}$-good and for any $J \in \mathcal{S}_\phi$ we have $J_+ \subseteq A(\phi, J)$. $\square$

The first part of the above proposition provides an estimate of the limitation for making the sharp inequality (1.3) an almost equality using dyadic step functions.

5. Bellman functions for the Maximal operator

Here we use the results of the previous sections to study the Bellman functions for $M_{\mathcal{T}}$ defined by (1.9) for any $p > 1$.

Let us fix $p > 1$ and a nonnegative $\phi \in L^p(X, \mu)$ such that $\int_X \phi^p \, d\mu = F$ and $\int_X \phi \, d\mu = f$ (where $f^p \leqslant F$). For any $I \in \mathcal{T}$ we apply Theorem 3 for ϕ restricted to I and for the tree $\mathcal{T}(I)$ on the probability space $(I, \frac{1}{\mu(I)}\mu)$ consisting of all elements of $\mathcal{T}$ contained in I to get

$$\mathrm{Av}_I[(M_{\mathcal{T}(I)}(\phi\chi_I))^p] \leqslant \mathrm{Av}_I(\phi^p)\omega_p\left(\frac{(\mathrm{Av}_I(\phi))^p}{\mathrm{Av}_I(\phi^p)}\right)^p. \tag{5.1}$$

Next fix $L \geqslant f$ and let $I_1, I_2, \ldots$ be all the maximal elements (if any) of $\{J \in \mathcal{T} : \mathrm{Av}_J(\phi) \geqslant L\}$. It is clear that the I_j's are pairwise almost disjoint and that writing $K = \bigcup_j I_j$

$$\max(M_{\mathcal{T}}\phi, L) = L\chi_{X \setminus K} + \sum_j (M_{\mathcal{T}(I_j)}(\phi\chi_{I_j}))^p \chi_{I_j}. \tag{5.2}$$

Therefore writing $k = \sum_j \mu(I_j)$ and using (5.1) for $I_1, I_2, \ldots$ we get

$$\int_X (\max(M_{\mathcal{T}}\phi, L))^p \, d\mu \leqslant L^p(1-k) + \sum\nolimits_j \alpha_j \omega_p\left(\frac{\beta_j}{\alpha_j}\right)^p \tag{5.3}$$

where

$$\alpha_j = \int_{I_j} \phi^p \, d\mu \geqslant \beta_j = \left(\frac{1}{\mu(I_j)^{1/q}} \int_{I_j} \phi \, d\mu\right)^p. \tag{5.4}$$

Now we write

$$A = \sum_j \alpha_j = \int_K \phi^p \, d\mu \leqslant F \quad \text{and} \quad B = \sum_j \left(\mu(I_j)^{p/q}\beta_j\right)^{1/p} = \int_K \phi \, d\mu \leqslant f \tag{5.5}$$

note that

$$\frac{B^p}{k^{p-1}} = \frac{\left(\sum_j (\mu(I_j)^{p-1}\beta_j)^{1/p}\right)^p}{\left(\sum_j \mu(I_j)\right)^{p-1}} \leqslant \sum_j \beta_j \leqslant A \tag{5.6}$$

and use the concavity of ω_p^p provided by Lemma 2(ii) to conclude that

$$\begin{aligned}\int_X (\max(M_{\mathcal{T}}\phi, L))^p \, d\mu &\leqslant L^p(1-k) + A\omega_p\left(\frac{\sum_j \beta_j}{A}\right)^p \\ &\leqslant L^p(1-k) + A\omega_p\left(\frac{B^p}{k^{p-1}A}\right).\end{aligned} \tag{5.7}$$

The parameters A, B and k satisfy the following inequalities:

$$\begin{gathered}\frac{B^p}{k^{p-1}} \leqslant A \leqslant F, \quad B \leqslant f, \quad 0 \leqslant k \leqslant 1 \quad \text{and} \\ (f-B)^p \leqslant (1-k)^{p-1}(F-A)\end{gathered} \tag{5.8}$$

the last one being just $(\int_{X\backslash K} \phi \, d\mu)^p \leqslant \mu(X\backslash K)^{p-1} \int_{X\backslash K} \phi^p \, d\mu$ (we also have $kB \geqslant L$ but we will not use it).

Conversely, assuming that $0<k<1$, $B<f$ and A, B satisfy the inequalities (5.8) we fix δ in $(0,1)$ we use Lemma 1 to pick a family $\{I_1, I_2, \dots\}$ of pairwise almost disjoint elements of $\mathcal{T}$ such that $\sum_j \mu(I_j) = k$ and since $\frac{B^p}{k^{p-1}} \leqslant A$ using Theorem 3 for each j we choose a nonnegative $\phi_j \in L^p(I_j, \frac{1}{\mu(I_j)}\mu)$ such that

$$\int_{I_j} \phi_j^p \, d\mu = \frac{A}{k}\mu(I_j), \quad \int_{I_j} \phi_j \, d\mu = \frac{B}{k}\mu(I_j) \tag{5.9}$$

and

$$\int_{I_j} (M_{\mathcal{T}(I_j)}(\phi_j))^p \, d\mu \geqslant \delta \frac{A}{k} \omega_p\left(\frac{B^p}{k^{p-1}A}\right)^p \mu(I_j). \tag{5.10}$$

Next, we choose a $\psi \in L^p(X\backslash K, \mu)$ such that

$$\int_{X\backslash K} \psi^p \, d\mu = F - A > 0 \quad \text{and} \quad \int_{X\backslash K} \psi \, d\mu = f - B > 0, \tag{5.11}$$

where $K = \bigcup_j I_j$ which is possible by (5.8) and define

$$\phi = \psi \chi_{X \setminus K} + \sum_j \phi_j \chi_{I_j}. \tag{5.12}$$

Then we have $\int_X \phi^p \, d\mu = F$, $\int_X \phi \, d\mu = f \leqslant L$ and

$$\int_X (\max(M_{\mathcal{T}}\phi, L))^p \, d\mu \geqslant L^p(1-k) + \delta A \omega_p \left(\frac{B^p}{k^{p-1}A} \right)^p. \tag{5.13}$$

Letting now $\delta \to 1^-$ we have proved the following:

$$\mathcal{B}_p^{\mathcal{T}}(F, f, L) = \sup\{L^p(1-k) + A\omega_p \left(\frac{B^p}{k^{p-1}A} \right)^p : \frac{B^p}{k^{p-1}} \leqslant A < F, B < f, 0 < k < 1$$
$$\text{and } (f-B)^p \leqslant (1-k)^{p-1}(F-A)\}. \tag{5.14}$$

Inequalities (5.8) for k, A and B imply that $\frac{(f-B)^p}{(1-k)^{p-1}} + \frac{B^p}{k^{p-1}} \leqslant F$ and $A \leqslant F - \frac{(f-B)^p}{(1-k)^{p-1}} = C < F$ and so using now Lemma 2(ii) we conclude that $A\omega_p(\frac{B^p}{k^{p-1}A})^p \leqslant C\omega_p(\frac{B^p}{k^{p-1}C})^p$ and hence by writing $x = \frac{f-B}{1-k}$ and $y = \frac{B}{k}$ we have proved the following.

Proposition 2. *For any tree $\mathcal{T}$ on (X, μ) and any $p > 1$ we have*

$$\mathcal{B}_p^{\mathcal{T}}(F, f, L)$$
$$= \sup\{L^p(1-k) + (F-(1-k)x^p)\omega_p \left(\frac{ky^p}{F-(1-k)x^p} \right)^p :$$
$$x, y > 0, 0 < k < 1, (1-k)x + ky = f \text{ and } (1-k)x^p + ky^p \leqslant F\}.$$

The above proposition will be exploited in the next section to prove Theorem 1.

6. Proof of Theorem 1

Here we will fix $p > 1$ and F, f, L with $f^p < F$ and $0 < f < L$ and consider the surface

$$S = \{(k, x, y) \in \mathbb{R}^3 : (1-k)x + ky = f\}$$

and the function

$$R(k, x, y) = L^p(1-k) + (F-(1-k)x^p)\omega_p \left(\frac{ky^p}{F-(1-k)x^p} \right)^p \tag{6.1}$$

on the subset

$$W = \{(k, x, y) \in S : x, y \geqslant 0, k \in [0, 1] \text{ and } (1-k)x^p + ky^p \leqslant F\} \tag{6.2}$$

of S. It is easy to see that S is a regular surface. Also R is continuously differentiable when $(1-k)x^p + ky^p < F$ and using Lemma 2(i) and the fact that $H_p(\omega_p(z)) = z$ for any $z \in [0,1]$ we can compute

$$\frac{\partial R}{\partial x} = (1-k)\frac{px^{p-1}}{p-1}\frac{\omega_p(Z)^p}{1-\omega_p(Z)},$$

$$\frac{\partial R}{\partial y} = k\frac{py^{p-1}}{p-1}\frac{\omega_p(Z)}{1-\omega_p(Z)},$$

$$\frac{\partial R}{\partial k} = -L^p + x^p\omega_p(Z)^p + \frac{1}{p-1}\frac{\omega_p(Z)}{1-\omega_p(Z)}(y^p - Zx^p), \tag{6.3}$$

where

$$Z = Z(k,x,y) = \frac{ky^p}{F-(1-k)x^p} < 1. \tag{6.4}$$

We want to compute the following quantity:

$$M = M(F,f,L) = \sup_{(k,x,y)\in W} R(k,x,y). \tag{6.5}$$

We have

Lemma 10. *R (as a function on S) has a critical point in the interior of W with respect to the surface S if and only if $L < \frac{p}{p-1}f$ in which case this critical point (k_1, x_1, y_1) is unique and satisfies*

$$R(k_1,x_1,y_1) = F\omega_p\left(\frac{pL^{p-1}f - (p-1)L^p}{F}\right)^p. \tag{6.6}$$

Proof. If R has such a critical point (k,x,y) then there would exist $\lambda \in \mathbb{R}$ such that $\frac{\partial R}{\partial x} = \lambda(1-k)$, $\frac{\partial R}{\partial y} = \lambda k$ and $\frac{\partial R}{\partial k} = \lambda(y-x)$ which using (6.3) give

$$y = \omega_p(Z)x = L \tag{6.7}$$

and so we must have $x = \frac{f-kL}{1-k}$, $k < \frac{f}{L}$ and $\frac{L}{x} = \omega_p(Z) < \frac{p}{p-1}$ which imply that $L < \frac{p}{p-1}f$. Therefore if $L \geqslant \frac{p}{p-1}f$ the function R has no critical point in the relative interior of W.

On the other hand if $L<\frac{p}{p-1}f$ then $\omega_p(Z)=\frac{L(1-k)}{f-kL}$ and so k must satisfy

$$-(p-1)\left(\frac{L(1-k)}{f-kL}\right)^p+p\left(\frac{L(1-k)}{f-kL}\right)^{p-1}=Z=\frac{kL^p}{F-\dfrac{(f-kL)^p}{(1-k)^{p-1}}}$$

which is equivalent to

$$\begin{aligned}F&\left[-(p-1)\left(\frac{L(1-k)}{f-kL}\right)^p+p\left(\frac{L(1-k)}{f-kL}\right)^{p-1}\right]\\&=kL^p+\frac{(f-kL)^p}{(1-k)^{p-1}}\left[-(p-1)\left(\frac{L(1-k)}{f-kL}\right)^p+p\left(\frac{L(1-k)}{f-kL}\right)^{p-1}\right]\\&=kL^p-(p-1)(1-k)L^p+p(f-kL)L^{p-1}=-(p-1)L^p+pfL^{p-1}\end{aligned}$$

and so to

$$(p-1)\left(\frac{L(1-k)}{f-kL}\right)^p-p\left(\frac{L(1-k)}{f-kL}\right)^{p-1}=\frac{(p-1)L^p-pfL^{p-1}}{F}. \tag{6.8}$$

which by Lemma 5 has a unique solution $k_1=k_1(F,f,L)$ which, in case $fL^{p-1}\geqslant F$, by Lemma 6 satisfies $0<k_1<k_0\leqslant\frac{f}{L}$ and so using Lemma 4 the point (k_1,x_1,y_1) where $y_1=L$ and $x_1=\frac{f-k_1L}{1-k_1}$ is always in the interior of W (with respect to S). Moreover $R(k_1,x_1,y_1)=L^p(1-k_1)+(F-(1-k_1)x_1^p)(\frac{L}{x_1})^p=F(\frac{L}{x_1})^p$ and by (6.8) we have $H_p(\frac{L}{x_1})=\frac{pfL^{p-1}-(p-1)L^p}{F}$. This completes the proof of this lemma. $\square$

Now choose a sequence $(k_m,x_m,y_m)\in W$, $m>1$ such that $R(k_m,x_m,y_m)\to M$ (note that W is not bounded). We may assume that k_m converges and consider the following cases:

Case 1: $k_m\to 0$. Then $k_my_m^p\leqslant F$ and so $k_my_m\to 0$ which implies that $(1-k_m)x_m\to f$. Since $\omega_p\leqslant\frac{p}{p-1}$, in this case we have

$$M=\lim_m R(k_m,x_m,y_m)\leqslant L^p+\left(\frac{p}{p-1}\right)^p(F-f^p)=R(0,f,0). \tag{6.9}$$

Case 2: $k_m\to 1$. Then $(1-k_m)x_m^p\leqslant F$ and so $(1-k_m)x_m\to 0$ which implies that $y_m\to f$. Also we may assume $(1-k_m)x_m^p\to z\leqslant F-f^p$ and therefore (using Lemma 2(iii))

$$M=\lim_m R(k_m,x_m,y_m)=(F-z)\omega_p\left(\frac{f^p}{F-z}\right)^p\leqslant F\omega_p\left(\frac{f^p}{F}\right)^p=R(1,0,f). \tag{6.10}$$

Case 3: $k_m\to\bar{k}\in(0,1)$. Here x_m and y_m are bounded and so we may assume that $x_m\to\bar{x}$, $y_m\to\bar{y}$ and $(\bar{k},\bar{x},\bar{y})\in W$ and so $M=\lim_m R(k_m,x_m,y_m)=R(\bar{k},\bar{x},\bar{y})$.

Applying Lemma 7 for this $\bar{k}$ and with $B = \bar{k}\bar{y} = f - (1-\bar{k})\bar{x}$ we conclude that $(1-\bar{k})\bar{x}^p + \bar{k}\bar{y}^p < F$ and $\bar{x}, \bar{y} > 0$ hence $(\bar{k}, \bar{x}, \bar{y})$ is in the relative interior of W and so it must be a critical point of R on S.

Now we can compute M.

Lemma 11. *If* $L \geqslant \frac{p}{p-1} f$ *then* $M = R(0, f, 0) = L^p + (\frac{p}{p-1})^p (F - f^p)$.

Proof. Here by Lemma 10 there are no interior critical points hence Case 3 cannot happen. Noting that

$$R(0, f, 0) = L^p + \left(\frac{p}{p-1}\right)^p (F - f^p) \geqslant \left(\frac{p}{p-1}\right)^p F > R(1, 0, f). \tag{6.11}$$

completes the proof. □

Now for the other case.

Lemma 12. *If* $L < \frac{p}{p-1} f$ *then* $M = R(k_1, x_1, y_1) = F\omega_p \left(\frac{pL^{p-1} f - (p-1)L^p}{F}\right)^p$.

Proof. Here the unique interior critical point (k_1, x_1, y_1) is provided by Lemma 10. We have:

Since, by Lemma 2(iv), $pL^{p-1} f - (p-1)L^p < f^p$ and ω_p is decreasing we have $R(1, 0, f) < R(k_1, x_1, y_1)$.

Now consider the function

$$\lambda(k) = R\left(k, \frac{f - kL}{1-k}, L\right) \tag{6.12}$$

defined for $k \geqslant 0$ small so that $(k, \frac{f-kL}{1-k}, L) \in W$ (note that $f^p < F$). We have $\lambda(0) = R(0, f, L) = R(0, f, 0)$ by the definition of R. Then we compute

$$\begin{aligned}\lambda'(0) &= \frac{\partial R}{\partial k}(0, f, L) + \frac{\partial R}{\partial x}(0, f, L)(f - L) \\ &= \frac{L^p}{p-1}\left[-(p-1)^2 z^p + p^2 z^{p-1} - (2p-1)\right] = \frac{L^p}{p-1}\tau(z),\end{aligned}$$

where $z = \frac{pf}{(p-1)L} \in (1, \frac{p}{p-1}]$ by our assumptions. But $\tau'(u) = p(p-1)[-(p-1)u + p]u^{p-2} > 0$ for all $u \in (1, \frac{p}{p-1})$ hence $\tau(z) > \tau(1) = 0$ which implies $\lambda'(0) > 0$. Hence $M > R(0, f, 0)$ which completes the proof. □

Proposition 2 combined with the above two lemmas complete the proof of Theorem 1.

7. The Bellman function for the Carleson imbedding

Here we study the Bellman function $\tilde{\mathcal{B}}_p^{\mathcal{T}}(F,f,k)$ defined in the introduction. First we will prove (1.17). First we need the following.

Lemma 13. *Suppose $\lambda_I \geqslant 0$ for each $I \in \mathcal{T}$, that $\lambda_I = 0$ for all but finitely many I's and that $\sum_{J\in\mathcal{T}\,:\,J\subseteq I}\lambda_J \leqslant \mu(I)$ for every $I\in\mathcal{T}$. Then we can choose pairwise disjoint measurable $A_I \subseteq X$ such that $A_I\subseteq I$ and $\lambda_I = \mu(A_I)$ for each $I\in\mathcal{T}$.*

Proof. Let $m>0$ be maximal with $\lambda_I>0$ for some $I\in\mathcal{T}_{(m)}$. Then since $\lambda_I\leqslant\mu(I)$ for all $I\in\mathcal{T}_{(m)}$ there exist pairwise disjoint A_I's with $\lambda_I=\mu(A_I)$ for each $I\in\mathcal{T}_{(m)}$. Assuming that we have chosen pairwise disjoint A_J's for all $J\in\bigcup_{j=0}^{s-1}\mathcal{T}_{(m-j)}$ with $\lambda_J=\mu(A_J)$ for each $I\in\mathcal{T}_{(m-s)}$ with $\lambda_I>0$ we note that

$$\lambda_I + \sum_{J\subseteq I, J\neq I}\mu(A_J)\leqslant\mu(I) \quad \text{so } \lambda_I\leqslant\mu\left(I\Big\backslash\bigcup_{J\subseteq I,J\neq I}A_J\right)$$

and so we may choose measurable $A_I\subseteq(I\backslash\bigcup_{J\subseteq I,J\neq I}A_J)\cap E(\mathcal{T})$ (see (2.1)) with $\mu(A_I)=\lambda_I$. Continuing this way and setting $A_I=\emptyset$ whenever $\lambda_I=0$ completes the proof. $\square$

Now we have the following.

Proposition 3. *$\tilde{\mathcal{B}}_p^{\mathcal{T}}(F,f,k)$ is equal to the supremum $\mathcal{D}_p^{\mathcal{T}}(F,f,k)$ of the quantity $\int_K(M_{\mathcal{T}}\phi)^p d\mu$ where $\phi\geqslant0, \phi\in L^p(X,\mu)$ is such that $\int_X\phi^p\,d\mu=F, \int_X\phi\,d\mu=f$ and K is a measurable subset of X with $\mu(K)=k$.*

Proof. Suppose first that ϕ,λ_I are as in the definition of $\tilde{\mathcal{B}}_p^{\mathcal{T}}(F,f,k)$ and that all but finitely many of the λ_I's are 0. Using the previous Lemma we pick the A_I's and since each A_I is contained in the corresponding I we have

$$\sum_{I\in\mathcal{T}}\lambda_I(\mathrm{Av}_I(\phi))^p\leqslant\int_X\sum_{I\in\mathcal{T}}(M_{\mathcal{T}}\phi)^p\chi_{A_I}\,d\mu=\int_{\bigcup_{I\in\mathcal{T}}A_I}(M_{\mathcal{T}}\phi)^p\,d\mu\leqslant\mathcal{D}_p^{\mathcal{T}}(F,f,k)$$

since $\mu(\bigcup_{I\in\mathcal{T}}A_I)=k$. In the general case for any finite subset $\mathcal{A}$ of $\{I\in\mathcal{T}:\lambda_I>0\}$ we have as before

$$\sum_{I\in\mathcal{A}}\lambda_I(\mathrm{Av}_I(\phi))^p\leqslant\mathcal{D}_p^{\mathcal{T}}\left(F,f,\sum_{I\in\mathcal{A}}\lambda_I\right)\leqslant\mathcal{D}_p^{\mathcal{T}}(F,f,k).$$

Hence $\tilde{\mathcal{B}}_p^{\mathcal{T}}(F,f,k)\leqslant\mathcal{D}_p^{\mathcal{T}}(F,f,k)$.

On the other hand, suppose that $\phi\geqslant0$ is such that $\int_X\phi^p\,d\mu=F, \int_X\phi\,d\mu=f$ and K is a measurable subset of X with $\mu(K)=k$. Fixing $\delta\in(0,1)$ we define for each

$x \in K$ the interval $I(x)$ that is maximal among all $I \in \mathcal{T}$ such that $x \in I$ and $\mathrm{Av}_I(\phi) > (1-\delta)M_{\mathcal{T}}\phi$ and then for each $I \in \mathcal{T}$ let $A_I = \{x \in K : I(x) = I\} \subseteq I$. It is easy to see that the A_I's are measurable and pairwise disjoint and that writing $\lambda_I = \mu(A_I)$ we have

$$\sum_{J \subseteq I} \lambda_J \leqslant \mu(I \cap K) \leqslant \mu(I) \quad \text{for each } I \in \mathcal{T}, \ \sum_{I \in \mathcal{T}} \lambda_I = \mu(X \cap K) = k \tag{7.1}$$

and so

$$(1-\delta)^p \int_K (M_{\mathcal{T}}\phi)^p \, d\mu \leqslant \sum_{I \in \mathcal{T}} \lambda_I (\mathrm{Av}_I(\phi))^p \leqslant \tilde{\mathcal{B}}_p^{\mathcal{T}}(F, f, k). \tag{7.2}$$

Letting $\delta \to 0^+$ completes the proof. $\square$

Next, suppose that ϕ, K satisfy the requirements in the definition of $\mathcal{D}_p^{\mathcal{T}}(F, f, k)$ and choose $u > 0$ such that

$$\mu(\{M_{\mathcal{T}}\phi > u\}) \leqslant k \leqslant \mu(\{M_{\mathcal{T}}\phi \geqslant u\}) \tag{7.3}$$

and choose a measurable D such that $V_1 = \{M_{\mathcal{T}}\phi > u\} \subseteq D \subseteq \{M_{\mathcal{T}}\phi \geqslant u\} = V_2$ and $\mu(D) = k$. Since $M_{\mathcal{T}}\phi \leqslant u$ on $K \backslash V_1$ it is easy to see that

$$\int_K (M_{\mathcal{T}}\phi)^p \, d\mu \leqslant \int_D (M_{\mathcal{T}}\phi)^p \, d\mu \tag{7.4}$$

and defining $s \in [0,1]$ by $\mu(D) = s\mu(V_1) + (1-s)\mu(V_2)$ we also have (since $M_{\mathcal{T}}\phi = u$ on $V_2 \backslash V_1$)

$$\int_D (M_{\mathcal{T}}\phi)^p \, d\mu = s \int_{V_1} (M_{\mathcal{T}}\phi)^p \, d\mu + (1-s) \int_{V_2} (M_{\mathcal{T}}\phi)^p \, d\mu. \tag{7.5}$$

Now since each of the V_1, V_2 is a union of families $\{I_j^{(1)}\}$, $\{I_r^{(2)}\}$ consisting of pairwise almost disjoint elements maximal under $\mathrm{Av}_I(\phi) > u$ (resp. $\geqslant u$) and we clearly have $M_{\mathcal{T}}\phi = M_{\mathcal{T}(I)}\phi$ for each of those I's, arguing as in Section 5 and using (7.4) and (7.5) we have

$$\int_K (M_{\mathcal{T}}\phi)^p \, d\mu \leqslant \sum_j s\alpha_j^{(1)} \omega_p \left(\frac{\beta_j^{(1)}}{\alpha_j^{(1)}} \right)^p + \sum_r (1-s)\alpha_r^{(2)} \omega_p \left(\frac{\beta_r^{(2)}}{\alpha_r^{(2)}} \right)^p, \tag{7.6}$$

where $\alpha_j^{(1)} = \int_{I_j^{(1)}} \phi^p \, d\mu \geqslant \beta_j^{(1)} = (\mu(I_j^{(1)})^{-1/q} \int_{I_j^{(1)}} \phi \, d\mu)^p$, $\alpha_r^{(2)} = \int_{I_r^{(2)}} \phi^p \, d\mu \geqslant \beta_r^{(2)} = (\mu(I_r^{(2)})^{-1/q} \int_{I_r^{(2)}} \phi \, d\mu)^p$. Hence using Lemma 2(ii) we have

$$\int_K (M_{\mathcal{T}}\phi)^p \, d\mu \leqslant A\omega_p \left(\frac{\sum_j s\beta_j^{(1)} + \sum_r (1-s)\beta_r^{(2)}}{A} \right)^p, \tag{7.7}$$

where

$$A = \sum_j s\alpha_j^{(1)} + \sum_r (1-s)\alpha_r^{(2)} = s\int_{V_1} \phi^p \, d\mu + (1-s)\int_{V_2} \phi^p d\mu \leqslant F. \quad (7.8)$$

Setting

$$B = s\int_{V_1} \phi \, d\mu + (1-s)\int_{V_2} \phi \, d\mu \leqslant f \quad (7.9)$$

and noting that $k = \mu(D) = \sum_j s\mu(I_j^{(1)}) + \sum_r (1-s)\mu(I_r^{(2)})$ it is easy to see using (4.19) that

$$\frac{B^p}{k^{p-1}} \leqslant \sum_j s\beta_j^{(1)} + \sum_r (1-s)\beta_r^{(2)} \leqslant A \quad (7.10)$$

and so since ω_p is decreasing we have

$$\int_K (M_{\mathcal{T}}\phi)^p \, d\mu \leqslant A\omega_p \left(\frac{B^p}{k^{p-1}A}\right)^p. \quad (7.11)$$

Moreover we note that A, B satisfy all the inequalities in (5.8) the last now being just

$$\left(\int_X \phi\eta \, d\mu\right)^p \leqslant \left(\int_X \eta \, d\mu\right)^{p-1} \left(\int_X \phi^p \eta \, d\mu\right) \quad (7.12)$$

where $\eta = s\chi_{X\setminus V_1} + (1-s)\chi_{X\setminus V_2}$.

Conversely for any A, B satisfying (5.8) (k is here fixed) the functions ϕ defined by (5.12) complete the proof of the following equality:

$$\mathcal{D}_p^{\mathcal{T}}(F, f, k) = \sup\{R_k(B) : 0 \leqslant B \leqslant f \quad \text{and} \quad h_k(B) \leqslant F\} \quad (7.13)$$

where R_k, h_k are defined by (3.11) and (3.12).

Hence Lemma 7 implies the following.

Theorem 5. *We have*

$$\mathcal{D}_p^{\mathcal{T}}(F, f, k) == \left[F\omega_{p,k}\left(\frac{f^p}{F}\right)^p - (1-k)f^p\right]\left[\frac{1-(1-k)\omega_{p,k}(\frac{f^p}{F})^{-1}}{k}\right]^p. \quad (7.14)$$

Now combining Theorem 5 with Proposition 3 it is easy to complete the proof of Theorem 2.

Acknowledgments

A.D.M. thanks the referee for his remarks and for suggesting Refs. [6–8].

References

[1] D.L. Burkholder, Boundary value problems and sharp inequalities for martingale transforms, Ann. Prob. 12 (1984) 647–702.

[2] D.L. Burkholder, Martingales and Fourier analysis in Banach spaces, C.I.M.E. Lectures Varenna Como, Italy, 1985, Lecture Notes Math. 1206 (1986) 61–108.

[3] D.L. Burkholder, Explorations in martingale theory and its applications, 'Ecole d'Et'e de Probabilit'es de Saint-Flour XIX-1989, Lecture Notes Math. 1464 (1991) 1–66.

[4] F. Nazarov, S. Treil, The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis, Algebra i Analyz 8 (5) (1996) 32–162.

[5] F. Nazarov, S. Treil, A. Volberg, The Bellman functions and two-weight inequalities for Haar multipliers, J. Amer. Math. Soc. 12 (4) (1999) 909–928.

[6] F. Nazarov, S. Treil, A. Volberg, Bellman function in stochastic optimal control and harmonic analysis (how our Bellman function got its name), Oper. Theory: Adv. Appl. 129 (2001) 393–424.

[7] L. Slavin, V. Vasyunin, Bellman function for the sharp classical and dyadic John-Nirenberg inequality, preprint.

[8] A. Volberg, Bellman approach to some problems in harmonic analysis, Seminaire des Equations aux derivées partielles, Ecole Polytéchnique (2002) 1–14.

[9] G. Wang, Sharp maximal inequalities for conditionally symmetric martingales and Brownian motion, Proc. Amer. Math. Soc. 112 (1991) 579–586.

Further reading

L. Grafakos, S. Montgomery-Smith, Best constants for uncentered maximal functions, Bull. London Math. Soc. 29 (1) (1997) 60–64.

M. de Guzmán, Real variable methods in Fourier analysis, North-Holland Mathematical Studies, Vol. 46, Notas Math. 75 (1981).

E.M. Stein, Harmonic Analysis, in: Princeton Mathematical Series, Vol. 43, Princeton University Press, Princeton, NA, 1993.